

Data Warehouse dan Integrasi Data

Mata Kuliah Business Intelligence – Memahami Penyimpanan dan Integrasi Data



Tujuan Pembelajaran



Konsep Data Warehouse

Memahami arsitektur dan karakteristik sistem data warehouse sebagai fondasi business intelligence



Proses ETL

Menguasai tahapan Extract, Transform, dan Load untuk integrasi data dari berbagai sumber



OLTP vs OLAP

Membedakan sistem transaksional dan analitik serta penerapannya dalam konteks bisnis

Apa itu Data Warehouse?

Data Warehouse (DW) adalah repositori terpusat yang menyimpan data terintegrasi dari berbagai sumber untuk mendukung pengambilan keputusan strategis. Berbeda dengan database operasional, data warehouse dirancang khusus untuk analisis dan pelaporan bisnis.

Sistem ini mengumpulkan data historis dalam jumlah besar, mengorganisasikannya berdasarkan subjek tertentu, dan menyediakan akses cepat untuk query analitik yang kompleks. Data warehouse menjadi fondasi utama dalam arsitektur business intelligence modern.

Manfaat Utama:

- Konsolidasi data dari sistem yang berbeda
- Mendukung analisis tren dan pola bisnis
- Meningkatkan kecepatan pengambilan keputusan
- Menyediakan single source of truth



Karakteristik Data Warehouse

William Inmon, yang dikenal sebagai "Bapak Data Warehousing", mendefinisikan empat karakteristik fundamental yang membedakan data warehouse dari sistem database lainnya. Karakteristik ini menjadi standar industri dalam merancang solusi data warehouse.

1

Subject-Oriented

Data diorganisir berdasarkan subjek bisnis utama seperti pelanggan, produk, atau penjualan, bukan berdasarkan aplikasi atau transaksi. Fokus pada analisis subjek spesifik memudahkan pemahaman konteks bisnis.

2

Integrated

Data dari berbagai sumber heterogen diintegrasikan dengan standar penamaan, format, dan struktur yang konsisten. Proses integrasi menghilangkan inkonsistensi dan memastikan kualitas data yang tinggi.

3

Time-Variant

Menyimpan data historis dalam rentang waktu yang panjang, biasanya 5-10 tahun. Setiap record memiliki timestamp yang memungkinkan analisis tren temporal dan perbandingan periode.

4

Non-Volatile

Data bersifat read-only dan tidak berubah setelah dimuat. Update dilakukan melalui proses batch terjadwal, bukan real-time, memastikan stabilitas untuk analisis konsisten.

Arsitektur Data Warehouse



Data Sources

Sistem operasional, database transaksional, file eksternal, dan sumber data lain yang menyediakan raw data



ETL Layer

Proses ekstraksi, transformasi, dan loading yang membersihkan dan mengintegrasikan data



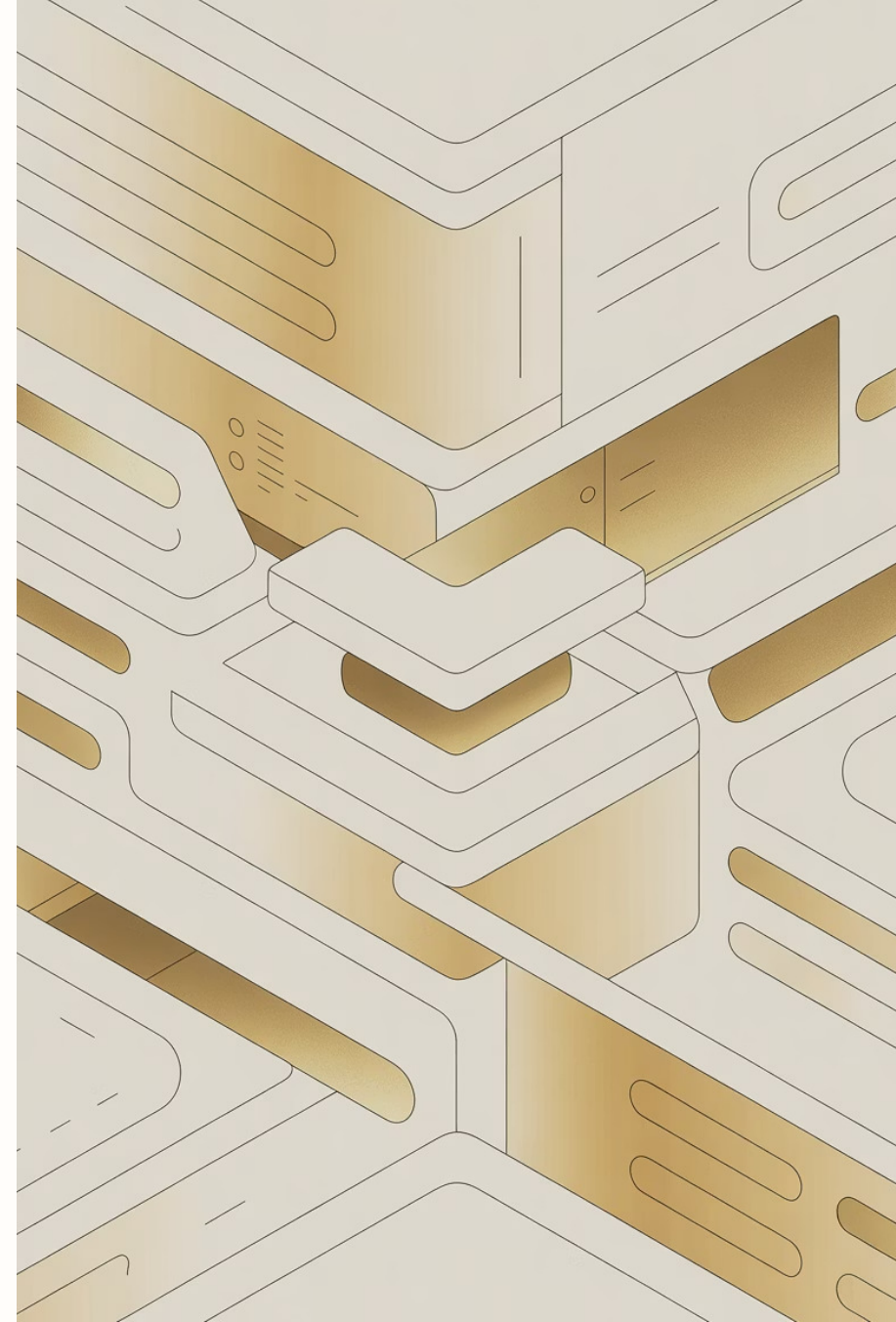
Data Storage

Repositori utama yang menyimpan data terintegrasi dalam struktur dimensional atau normalized



Presentation Layer

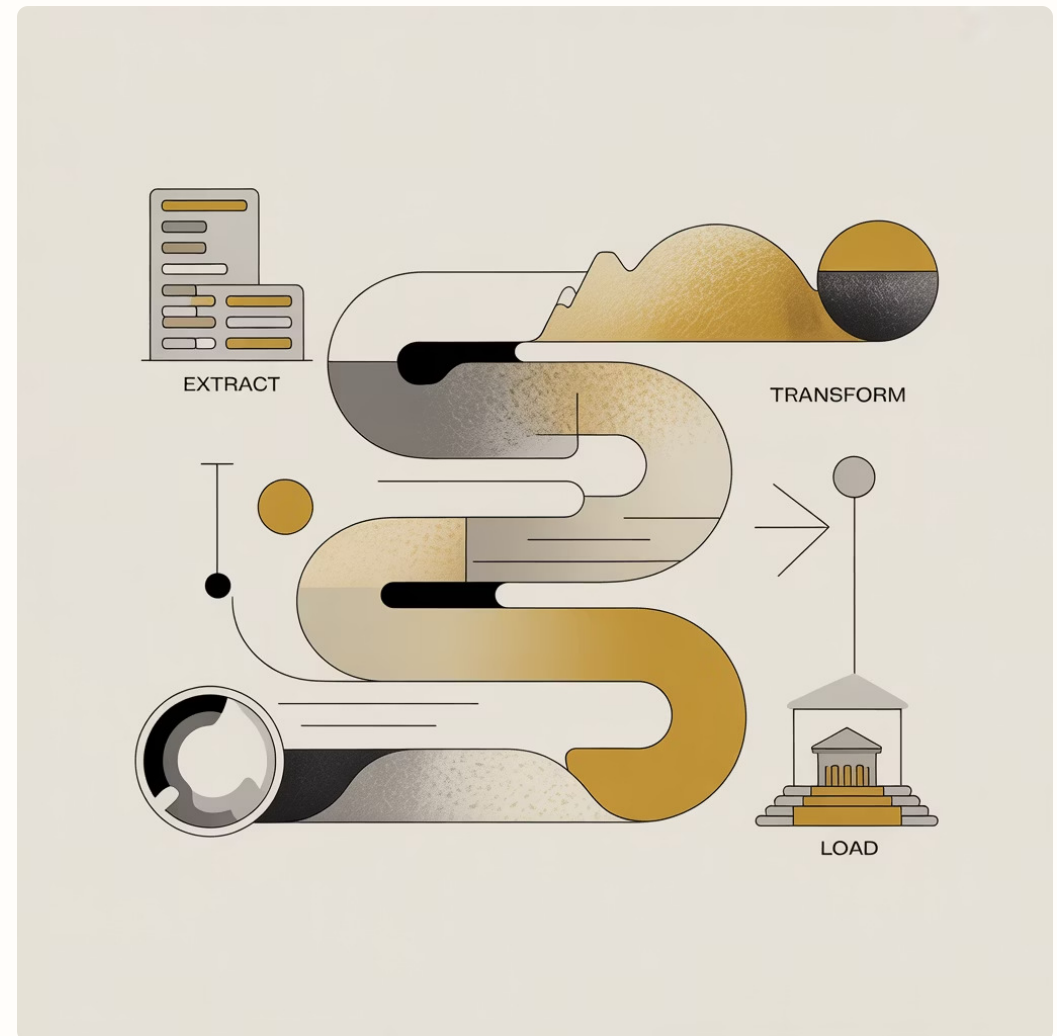
Tools analitik, dashboard, dan reporting yang mengakses data untuk end-users



Proses ETL: Extract, Transform, Load

ETL adalah proses kritis yang menggerakkan data dari sistem sumber ke data warehouse. Proses ini memastikan data yang masuk berkualitas tinggi, konsisten, dan siap untuk analisis.

Setiap tahap ETL memiliki peran spesifik dalam menjaga integritas dan nilai data. Proses ini biasanya berjalan secara terjadwal, seringkali pada malam hari untuk meminimalkan dampak terhadap sistem operasional.



01

Extract (Ekstraksi)

Mengambil data dari berbagai sumber seperti database OLTP, file CSV, API, atau sistem legacy

02

Transform (Transformasi)

Membersihkan, validasi, konversi format, agregasi, dan menerapkan business rules pada data

03

Load (Pemuatan)

Memuat data yang sudah ditransformasi ke dalam data warehouse dengan strategi full atau incremental

Tahap Extract: Ekstraksi Data

Full Extraction

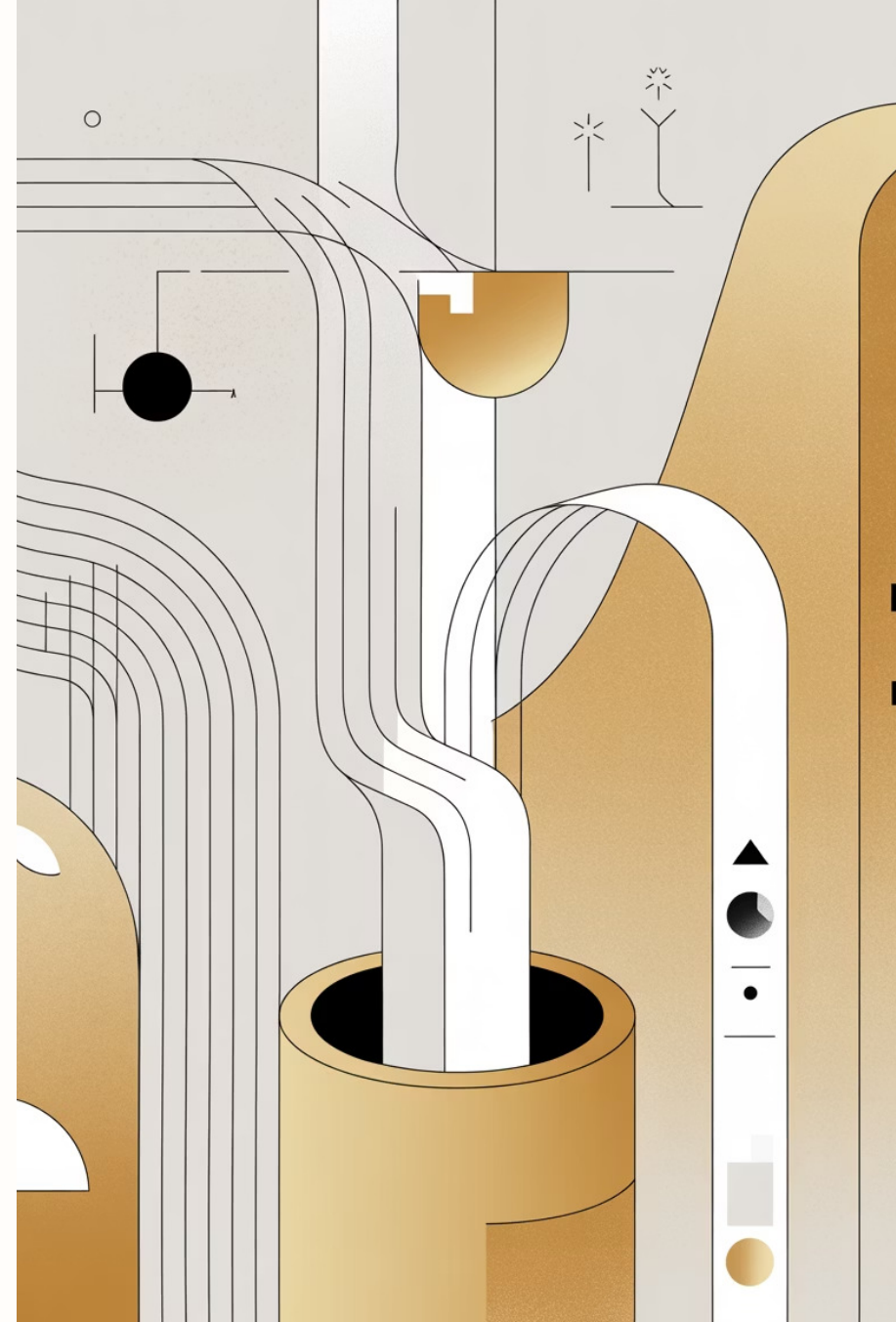
Mengambil seluruh data dari sistem sumber. Metode ini sederhana namun resource-intensive, cocok untuk initial load atau sistem dengan data volume kecil.

- Proses lebih lambat
- Beban tinggi pada sistem sumber
- Memastikan konsistensi penuh

Incremental Extraction

Hanya mengambil data yang berubah sejak ekstraksi terakhir. Lebih efisien dan mengurangi beban sistem, menjadi pilihan utama untuk operasi rutin.

- Proses lebih cepat
- Memerlukan mekanisme change detection
- Optimalisasi resource



Tahap Transform: Transformasi Data

Transformasi adalah tahap paling kompleks dalam ETL, di mana data mentah diubah menjadi informasi berkualitas tinggi yang siap dianalisis. Proses ini menerapkan berbagai aturan bisnis dan teknik data cleansing.



Data Cleansing

Menghilangkan duplikasi, memperbaiki kesalahan, mengisi nilai yang hilang, dan standardisasi format data untuk konsistensi



Data Integration

Menggabungkan data dari sumber berbeda, menyelesaikan konflik skema, dan memastikan referential integrity antar tabel



Aggregation

Menghitung summary metrics, membuat rollup data, dan menghasilkan derived attributes untuk mempercepat query analitik



Business Rules

Menerapkan logika bisnis spesifik seperti kalkulasi KPI, kategorisasi pelanggan, atau penentuan status berdasarkan kriteria

OLTP vs OLAP

Memahami perbedaan fundamental antara sistem Online Transaction Processing (OLTP) dan Online Analytical Processing (OLAP) sangat penting dalam merancang arsitektur data yang efektif. Kedua sistem ini dirancang untuk tujuan yang berbeda dan memiliki karakteristik teknis yang kontras.

OLTP - Transactional

Sistem operasional harian yang menangani transaksi bisnis real-time dengan volume tinggi. Fokus pada kecepatan, konsistensi, dan integritas data untuk operasi insert, update, dan delete.

OLAP - Analytical

Sistem analitik yang dirancang untuk query kompleks dan analisis multidimensional. Fokus pada kecepatan pembacaan data historis untuk mendukung pengambilan keputusan strategis.

OLTP Online Transaction Processing

Online transaction processing (OLTP) is a type of database management system (DBMS) that is designed to handle a large number of small, simple transactions. It is typically used for applications that require high performance and low latency, such as e-commerce, banking, and airline reservations.



OLAP Online Analytical Processing

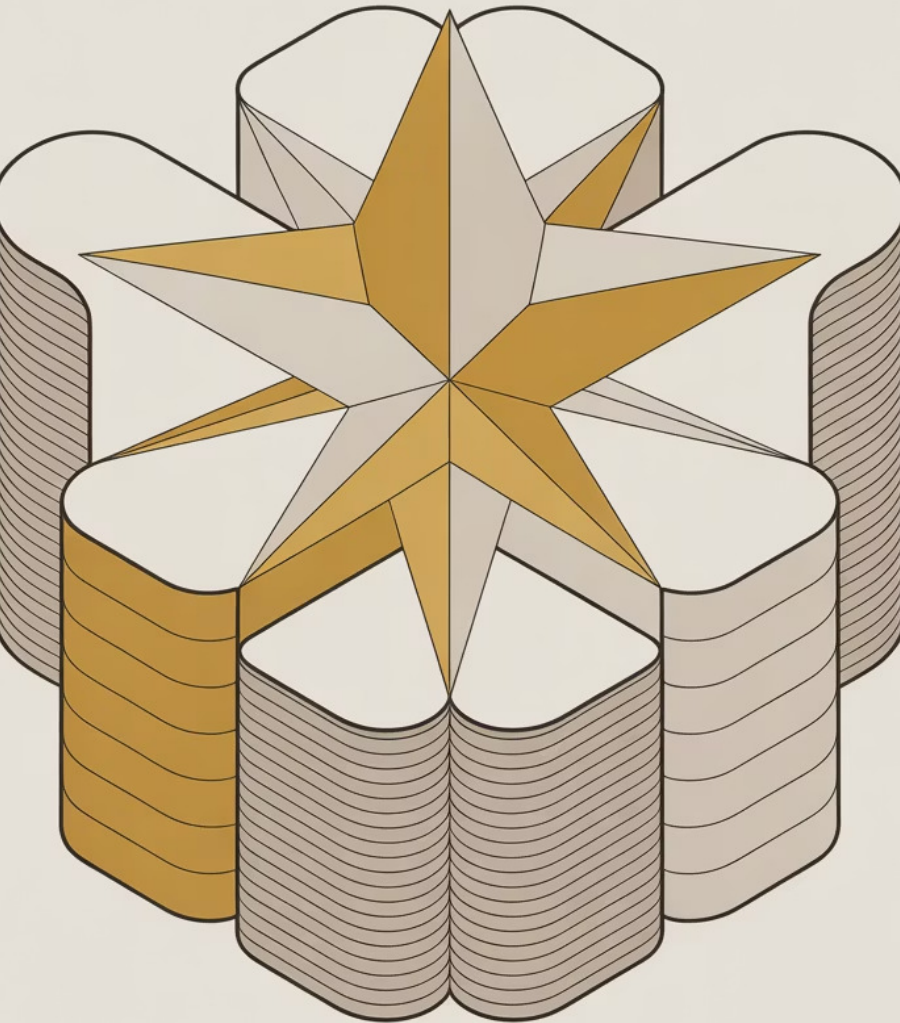
Online analytical processing (OLAP) is a type of database management system (DBMS) that is designed to handle complex, multi-dimensional queries. It is typically used for applications that require high performance and low latency, such as data warehousing, business intelligence, and financial reporting.



Perbandingan Detail: OLTP vs OLAP

Aspek	OLTP	OLAP
Tujuan	Mengelola transaksi operasional harian	Mendukung analisis dan pengambilan keputusan
Jenis Query	Simple, predefined, cepat (SELECT, INSERT, UPDATE)	Complex, ad-hoc, agregasi besar
Volume Data	Current data, GB hingga beberapa TB	Historical data, TB hingga PB
Desain Database	Normalized (3NF) untuk minimalisasi redundansi	Denormalized (Star/Snowflake) untuk optimasi query
Response Time	Milliseconds untuk single transaction	Seconds hingga minutes untuk analisis kompleks
Update Frequency	Real-time, continuous updates	Batch, periodic (harian/mingguan)
Users	Ribuan concurrent operational users	Puluhan analysts dan decision makers
Contoh Aplikasi	E-commerce checkout, banking transactions, inventory management	Sales forecasting, customer segmentation, trend analysis

Dimensi dan Fakta dalam Data Warehouse



Fact Table (Tabel Fakta)

Menyimpan data kuantitatif atau metrik bisnis yang dapat diukur dan diagregasi. Fact table berisi foreign keys yang menghubungkan ke dimension tables dan measures numerik.

Contoh measures:

- Total penjualan (revenue)
- Quantity produk terjual
- Profit margin
- Biaya operasional

Fact table biasanya sangat besar karena menyimpan detail setiap transaksi atau event bisnis.

Dimension Table (Tabel Dimensi)

Menyimpan atribut deskriptif yang memberikan konteks untuk analisis fakta. Dimension tables relatif kecil namun berisi banyak kolom deskriptif.

Contoh dimensi:

- Dimensi Waktu (tanggal, bulan, kuartal, tahun)
- Dimensi Produk (nama, kategori, brand)
- Dimensi Pelanggan (nama, lokasi, segment)
- Dimensi Lokasi (kota, provinsi, region)

Dimension tables memungkinkan slicing, dicing, dan drill-down dalam analisis OLAP.

Ringkasan & Kesimpulan

Data Warehouse sebagai Fondasi BI

Data warehouse adalah repositori terpusat yang mengintegrasikan data dari berbagai sumber, dirancang khusus untuk analisis dengan karakteristik subject-oriented, integrated, time-variant, dan non-volatile

ETL: Jembatan Data ke Insight

Proses ETL (Extract, Transform, Load) memastikan data berkualitas tinggi masuk ke warehouse melalui ekstraksi dari sumber, transformasi dengan business rules, dan loading ke target sistem

OLTP untuk Operasi, OLAP untuk Analisis

OLTP menangani transaksi operasional real-time dengan database normalized, sementara OLAP mendukung analisis kompleks menggunakan struktur dimensional untuk query performa tinggi

Poin Penting: Pemahaman konsep data warehouse, ETL, dan perbedaan OLTP-OLAP adalah fundamental untuk merancang solusi Business Intelligence yang efektif dan scalable dalam organisasi modern.